Docket No. 250389US6

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Kenichiro KOBAYASHI GAU:

SERIAL NO: New Application EXAMINER:

FILED: Herewith

FOR: SINGING VOICE SYNTHESIZING METHOD AND APPARATUS, PROGRAM, RECORDING MEDIUM AND ROBOT APPARATUS

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of **35 U.S.C. §120**.

☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of **35 U.S.C. §119(e)**: **Application No.** **Date Filed**

■ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of **35 U.S.C. §119**, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

COUNTRY	APPLICATION NUMBER	MONTH/DAY/YEAR
Japan	2003-079151	March 20, 2003

Certified copies of the corresponding Convention Application(s)

■ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

☐ were filed in prior application Serial No. filed

☐ were submitted to the International Bureau in PCT Application Number
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and

☐ (B) Application Serial No.(s)

☐ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Bradley D. Lytle
Registration No. 40,073

C. Irvin McClelland
Registration Number 21,124

Customer Number
22850
Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 05/03)

S04P0354US00

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年　３月２０日

出願番号
Application Number: 特願２００３－０７９１５１

[ST. 10/C]: [ＪＰ２００３－０７９１５１]

出願人
Applicant(s): ソニー株式会社

２００４年　１月１３日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫

出証番号　出証特２００３－３１１０７２３

【書類名】　特許願
【整理番号】　0390001805
【提出日】　平成15年 3月20日
【あて先】　特許庁長官　殿
【国際特許分類】　G06F 17/20
G10L 13/00
G10H 7/00
【発明者】
【住所又は居所】　東京都品川区北品川６丁目７番３５号　ソニー株式会社内
【氏名】　小林　賢一郎
【特許出願人】
【識別番号】　000002185
【氏名又は名称】　ソニー株式会社
【代理人】
【識別番号】　100067736
【弁理士】
【氏名又は名称】　小池　晃
【選任した代理人】
【識別番号】　100086335
【弁理士】
【氏名又は名称】　田村　榮一
【選任した代理人】
【識別番号】　100096677
【弁理士】
【氏名又は名称】　伊賀　誠司
【手数料の表示】
【予納台帳番号】　019530
【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9707387

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　歌声合成方法、歌声合成装置、プログラム及び記録媒体、並びにロボット装置

【特許請求の範囲】

【請求項１】　演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、

解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とする歌声合成方法。

【請求項２】　上記演奏データはＭＩＤＩファイルの演奏データであることを特徴とする請求項１記載の歌声合成方法。

【請求項３】　上記楽音生成工程は上記歌声の対象とした演奏データに係る楽音をミュートすることを特徴とする請求項１記載の歌声合成方法。

【請求項４】　上記楽音生成工程は上記演奏データのうち、予め指定されたトラックの演奏データに係る楽音をミュートすることを特徴とする請求項２記載の歌声合成方法。

【請求項５】　上記楽音生成工程は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することを特徴とする請求項１記載の歌声合成方法。

【請求項６】　上記歌声と上記楽音の同期を取ってミキシングするミキシング工程をさらに有することを特徴とする請求項１記載の歌声合成方法。

【請求項７】　上記ミキシング工程は上記歌声生成工程からの上記歌声と上記楽音生成工程からの上記楽音をミキシングする際に、それぞれの波形データを予め作成し重ね合わせることによりミキシングをすることを特徴とする請求項６記載の歌声合成方法。

【請求項８】　演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、

解析された音楽情報に基づき、歌声を生成する歌声生成手段と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手

段とを有することを特徴とする歌声合成装置。

【請求項９】　上記演奏データはＭＩＤＩファイルの演奏データであることを特徴とする請求項８記載の歌声合成装置。

【請求項１０】　上記楽音生成手段は上記歌声の対象とした演奏データに係る楽音をミュートすることを特徴とする請求項８記載の歌声合成装置。

【請求項１１】　上記楽音生成手段は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することを特徴とする請求項８記載の歌声合成装置。

【請求項１２】　上記歌声と上記楽音の同期を取ってミキシングするミキシング手段をさらに有することを特徴とする請求項８記載の歌声合成装置。

【請求項１３】　所定の処理をコンピュータに実行させるためのプログラムであって、

演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、

解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とするプログラム。

【請求項１４】　上記演奏データはＭＩＤＩファイルの演奏データであることを特徴とする請求項１３記載のプログラム。

【請求項１５】　上記歌声と上記楽音の同期を取ってミキシングするミキシング工程をさらに有することを特徴とする請求項１３記載のプログラム。

【請求項１６】　所定の処理をコンピュータに実行させるためのプログラムが記録されたコンピュータ読み取り可能な記録媒体であって、

演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、

解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とするプログラムが記録された記録媒体。

【請求項１７】　上記演奏データはＭＩＤＩファイルの演奏データであることを特徴とする請求項１６記載の記録媒体。

【請求項１８】　上記歌声と上記楽音の同期を取ってミキシングするミキシン

グ工程をさらに有することを特徴とする請求項１６記載の記録媒体。

【請求項１９】　供給された入力情報に基づいて動作を行う自律型のロボット装置であって、

入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、

解析された音楽情報に基づき、歌声を生成する歌声生成手段と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とするロボット装置。

【請求項２０】　上記演奏データはＭＩＤＩファイルの演奏データであることを特徴とする請求項１９記載のロボット装置。

【請求項２１】　上記歌声と上記楽音の同期を取ってミキシングするミキシング手段をさらに有することを特徴とする請求項２０記載のロボット装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、演奏データから歌声を合成する歌声合成方法、歌声合成装置、プログラム及び記録媒体、並びにロボット装置に関する。

【０００２】

【従来の技術】

コンピュータ等により、与えられた歌唱データから歌声を生成する技術は特許文献１に代表されるように既に知られている。

【０００３】

ＭＩＤＩ（musical instrument digital interface）データは代表的な演奏データであり、事実上の業界標準である。代表的には、ＭＩＤＩデータはＭＩＤＩ音源と呼ばれるデジタル音源（コンピュータ音源や電子楽器音源等のＭＩＤＩデータにより動作する音源）を制御して楽音を生成するのに使用される。ＭＩＤＩファイル（例えば、ＳＭＦ（standard ＭＩＤＩ file））には歌詞データを入れることができ、歌詞付きの楽譜の自動作成に利用される。

【０００４】

また、ＭＩＤＩデータを歌声又は歌声を構成する音素セグメントのパラメータ表現（特殊データ表現）として利用する試みも特許文献２に代表されるように提案されている。

【０００５】

しかし、これらの従来の技術においてはＭＩＤＩデータのデータ形式の中で歌声を表現しようとしているが、あくまでも楽器をコントロールする感覚でのコントロールであり、ＭＩＤＩ本来が持っている歌詞データを利用するものではなかった。

【０００６】

また、ほかの楽器用に作成されたＭＩＤＩデータを、修正を加えることなく歌声にすることはできなかった。

【０００７】

また、電子メールやホームページを読み上げる音声合成ソフトはソニー(株)の「Simple　Speech」をはじめ多くのメーカーから発売されているが、読み上げ方は普通の文章を読み上げるのと同じような口調であった。

【０００８】

ところで、電気的又は磁気的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、１９６０年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

【０００９】

最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面において、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬、猫のように４足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、あるいは、２足直立歩行を行う人間等の身体メカニズムや動作をモデルにしてデザインされた

「人間型」又は「人間形」ロボット（Humanoid Robot）等のロボット装置は、既に実用化されつつある。

【００１０】

これらのロボット装置は、産業用ロボットと比較して、エンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。また、そのようなロボット装置には、外部からの情報や内部の状態に応じて自律的に動作するものがある。

【００１１】

この自律的に動作するロボット装置に用いられる人工知能（ＡＩ：artificial intelligence）は、推論・判断等の知的な機能を人工的に実現したものであり、さらに感情や本能等の機能をも人工的に実現することが試みられている。このような人工知能の外部への表現手段としての視覚的な表現手段や自然言語の表現手段等のうちで、自然言語表現機能の一例として、音声を用いることが挙げられる。

【００１２】

【特許文献１】

特許第３２３３０３６号公報

【特許文献２】

特開平１１－９５７９８号公報

【００１３】

【発明が解決しようとする課題】

以上のように従来の歌声合成は特殊な形式のデータを用いていたり、仮にＭＩＤＩデータを用いていてもその中に埋め込まれている歌詞データを有効に活用できなかったり、ほかの楽器用に作成されたＭＩＤＩデータを歌い上げたりすることはできなかった。

【００１４】

本発明は、このような従来の実情に鑑みて提案されたものであり、例えばＭＩＤＩデータのような演奏データを活用して歌声を合成することが可能な歌声合成方法及び装置を提供することを目的とする。

【００１５】

さらに、本発明の目的は、ＭＩＤＩデータのような演奏データを活用する際、歌声に使用し、かつ歌声と共にもとの演奏データから楽音も再生可能した歌声合成方法及び装置を提供することである。

【００１６】

さらに、本発明の目的は、このような歌声合成機能をコンピュータに実施させるプログラム及び記録媒体を提供することである。

【００１７】

さらに、本発明の目的は、このような歌声合成機能を実現するロボット装置を提供することである。

【００１８】

【課題を解決するための手段】

本発明に係る歌声合成方法は、上記目的を達成するため、演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、解析された音楽情報に基づき、歌声を生成する歌声生成工程と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とする。

【００１９】

また、本発明に係る歌声合成装置は、上記目的を達成するため、演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、解析された音楽情報に基づき、歌声を生成する歌声生成手段と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とする。

【００２０】

この構成によれば、本発明に係る歌声合成方法及び装置は、演奏データを解析してそれから得られる歌詞や音の高さ、長さ、強さをもとにした音符情報に基づき歌声情報を生成し、その歌声情報をもとに歌声の生成を行うことができる。さらに、演奏データから歌声の伴奏としての楽音を再生することにより、伴奏のもとで歌詞を歌い上げることができる。

【００２１】

上記演奏データはＭＩＤＩファイル（例えばＳＭＦ）の演奏データであること

が好ましい。

【００２２】

上記楽音生成工程又は手段は歌声を目立たせるために上記歌声の対象とした演奏データに係る楽音をミュートする（楽音として出力しない）ことが好ましい。

【００２３】

あるいは、上記楽音生成工程又は手段は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することにより、カラオケ等におけるメロディガイド機能を果たすことができる。

【００２４】

また、上記楽音生成工程又は手段は上記ＭＩＤＩファイルの演奏データのうち、歌詞の対象等として予め指定されたトラックの演奏データに係る楽音をミュートすることが好ましい。

【００２５】

さらに、上記歌声と上記楽音の同期を取ってミキシングするミキシング工程又は手段を設けることが好ましい。ミキシングの方式としては、歌声と楽音のそれぞれの波形データを予め作成し重ね合わせることによりミキシングをすることとし、ミキシング結果を保存してもよい。

【００２６】

また、本発明に係るプログラムは、本発明の歌声合成機能をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【００２７】

さらに、本発明に係るロボット装置は、上記目的を達成するため、供給された入力情報に基づいて動作を行う自律型のロボット装置であって、入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、解析された音楽情報に基づき、歌声を生成する歌声生成手段と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とする。これにより、ロボットの持っているエンターテインメント性を格段に向上させることができる。

【００２８】

【発明の実施の形態】

以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【００２９】

先ず、本実施の形態における歌声合成装置の概略システム構成を図１に示す。ここで、この歌声合成装置は、少なくとも感情モデル、音声合成手段及び発音手段を有する例えばロボット装置に適用することを想定しているが、これに限定されず、各種ロボット装置や、ロボット以外の各種コンピュータＡＩ（artificial intelligence）等への適用も可能であることは勿論である。

【００３０】

図１において、ＭＩＤＩデータに代表される演奏データ１を解析する演奏データ解析部２は入力された演奏データ１を解析し演奏データ内にあるトラックやチャンネルの音の高さや長さ、強さを表す楽譜情報４に変換する。

【００３１】

図２に楽譜情報４に変換された演奏データ（ＭＩＤＩデータ）の例を示す。図２において、トラック毎、チャンネル毎にイベントが書かれている。イベントにはノートイベントとコントロールイベントが含まれる。ノートイベントは発生時刻（図中の時間の欄）、高さ、長さ、強さ（velocity）の情報を持つ。したがって、ノートイベントのシーケンスにより音符列又は音列が定義される。コントロールイベントは発生時刻、コントロールのタイプデータ（例えばビブラート、演奏ダイナミクス表現（expression））及びコントロールのコンテンツを示すデータを持つ。例えば、ビブラートの場合、コントロールのコンテンツとして、音の振れの大きさを指示する「深さ」、音の揺れの周期を指示する「幅」、音の揺れの開始タイミング（発音タイミングからの遅れ時間）を指示する「遅れ」の項目を有する。特定のトラック、チャンネルに対するコントロールイベントはそのコントロールタイプについて新たなコントロールイベント（コントロールチェンジ）が発生しない限り、そのトラック、チャンネルの音符列の楽音再生に適用される。さらに、ＭＩＤＩファイルの演奏データにはトラック単位で歌詞を記入する

ことができる。図２において、上方に示す「あるうひ」はトラック１に記入された歌詞の一部であり、下方に示す「あるうひ」はトラック２に記入された歌詞の一部である。すなわち図２の例は、解析した音楽情報（楽譜情報）の中に歌詞が埋め込まれた例である。

【００３２】

なお、図２において、時間は「小節：拍：ティック数」で表され、長さは「ティック数」で表され、強さは「０－１２７」の数値で表され、高さは４４０Hzが「Ａ４」で表される。また、ビブラートは、深さ、幅、遅れがそれぞれ「０－６４－１２７」の数値で表される。

【００３３】

図１に戻り、変換された楽譜情報４は歌詞付与部５に渡される。歌詞付与部５では楽譜情報４をもとに音符に対応した音の長さ、高さ、強さ、表情などの情報とともにその音に対する歌詞が付与された歌声情報６の生成を行う。

【００３４】

図３に歌声情報６の例を示す。図３において、「￥ｓｏｎｇ￥」は歌詞情報の開始を示すタグである。タグ「￥ＰＰ，Ｔ１０６７３０７５￥」は１０６７３０７５μsecの休みを示し、タグ「￥ｔｄｙｎａ　１１０　６４９０７５￥」は先頭から１０６７３０７５μsecの全体の強さを示し、タグ「￥ｆｉｎｅ１００￥」はＭＩＤＩのファインチューンに相当する高さの微調整を示し、タグ「￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｄｅｐ＝６４￥」、［￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｄｅｌ＝５０￥］、「￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｒａｔ＝６４￥」はそれぞれ、ビブラートの深さ、遅れ、幅を示す。また、タグ「￥ｄｙｎａ　１００￥」は音毎の強弱を示し、タグ「￥Ｇ４，Ｔ２８８４６１￥あ」はＧ４の高さで、長さが２８８４６１μsecの歌詞「あ」を示す。図３の歌声情報は図２に示す楽譜情報（ＭＩＤＩデータの解析結果）から得られたものである。

【００３５】

図３と図２の比較から分かるように、楽器制御用の演奏データ（例えば音符情報）が歌声情報の生成において十分に活用されている。例えば、歌詞「あるうひ」の構成要素「あ」について、「あ」以外の歌唱属性である「あ」の音の発生時

刻、長さ、高さ、強さ等について、楽譜情報（図２）中のコントロール情報やノートイベント情報に含まれる発生時刻、長さ、高さ、強さ等が直接的に利用され、次の歌詞要素「る」についても楽譜情報中の同じトラック、チャンネルにおける次のノートイベント情報が直接的に利用され、以下同様である。

【００３６】

図１に戻り、歌声情報６は歌声生成部７に渡される。歌声生成部７は音声合成器（speech synthesizer）を構成する。歌声生成部７においては歌声情報６をもとに歌声波形８の生成を行う。ここで、歌声情報６から歌声波形８を生成する歌声生成部７は例えば図４に示すように構成される。

【００３７】

図４において、歌声韻律生成部７－１は歌声情報６を歌声韻律データに変換する。波形生成部７－２は歌声韻律データを歌声波形８に変換する。

【００３８】

具体例として、「Ａ４」の高さの歌詞要素「ら」を一定時間伸ばす場合について説明する。ビブラートをかけない場合の歌声韻律データは、以下の表のように表される。

【００３９】

【表１】

[LABEL]	[PITCH]	[VOLUME]
0 ra	0 50	0 66
1000 aa		39600 57
39600 aa		40100 48
40100 aa		40600 39
40600 aa		41100 30
41100 aa		41600 21
41600 aa		42100 12
42100 aa		42600 3
42600 aa		
43100 a.		

【００４０】

この表において、[LABEL]は、各音韻の継続時間長を表したものである。すなわち、「ｒａ」という音韻（音素セグメント）は、０サンプルから１０００サンプルまでの１０００サンプルの継続時間長であり、「ｒａ」に続く最初の「ａａ」という音韻は、１０００サンプルから３９６００サンプルまでの３８６００サンプルの継続時間長である。また、[ＰＩＴＣＨ]は、ピッチ周期を点ピッチで表したものである。すなわち、０サンプル点におけるピッチ周期は５６サンプルである。ここでは「ら」の高さを変えないので全てのサンプルに渡り５６サンプルのピッチ周期が適用される。また、[VOLUME]は、各サンプル点での相対的な音量を表したものである。すなわち、デフォルト値を１００％としたときに、０サンプル点では６６％の音量であり、３９６００サンプル点では５７％の音量である。以下同様にして、４０１００サンプル点では４８％の音量等が続き４２６００サンプル点では３％の音量となる。これにより「ら」の音声が時間の経過と共に減衰することが実現される。

【００４１】

これに対して、ビブラートをかける場合には、例えば、以下に示すような歌声

韻律データが作成される。

【００４２】

【表２】

[LABEL]		[PITCH]		[VOLUME]	
0	ra	0	50	0	66
1000	aa	1000	50	39600	57
11000	aa	2000	53	40100	48
21000	aa	4009	47	40600	39
31000	aa	6009	53	41100	30
39600	aa	8010	47	41600	21
40100	aa	10010	53	42100	12
40600	aa	12011	47	42600	3
41100	aa	14011	53		
41600	aa	16022	47		
42100	aa	18022	53		
42600	aa	20031	47		
43100	a.	22031	53		
		24042	47		
		26042	53		
		28045	47		
		30045	53		
		32051	47		
		34051	53		
		36062	47		
		38062	53		
		40074	47		
		42074	53		
		43100	50		

【００４３】

この表の［ＰＩＴＣＨ］の欄に示すように、０サンプル点と１０００サンプル点におけるピッチ周期は５０サンプルで同じであり、この間は音声の高さに変化がないが、それ以降は、２０００サンプル点で５３サンプルのピッチ周期、４００９サンプル点で４７サンプルのピッチ周期、６００９サンプル点で５３のピッチ周期というようにピッチ周期が約４０００サンプルの周期（幅）を以て上下（５０±３）に振れている。これにより音声の高さの揺れであるビブラートが実現される。この［ＰＩＴＣＨ］の欄のデータは歌声情報６における対応歌声要素（例えば「ら」）に関する情報、特にノートナンバー（例えばＡ４）とビブラートコントロールデータ（例えば、タグ「￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｄｅｐ＝６４￥」、［￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｄｅｌ＝５０￥］、「￥ｖｉｂｒａｔｏ　ＮＲＰＮ＿ｒａｔ＝６４￥」）に基づいて生成される。

【００４４】

波形生成部７－２はこのような歌声音韻データに基づき、音素セグメントデータを記憶するデータメモリ（図示せず）から該当するサンプルを読み出して歌声波形８を生成する。すなわち、波形生成部７－２は、データメモリを参照しながら、歌声韻律データに示される音韻系列、ピッチ周期、音量等をもとに、なるべくこれに近い音素セグメントデータを検索してその部分を切り出して並べ、音声波形データを生成する。すなわち、データメモリには、例えば、ＣＶ（Consonant，Vowel）や、ＶＣＶ、ＣＶＣ等の形で音素セグメントデータが記憶されており、波形生成部７－２は、歌声韻律データに基づいて、必要な音素セグメントデータを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、歌声波形８を生成する。なお、歌声情報６から歌声波形８を生成する歌声生成部７については上記の例に限らず、任意の適当な公知の音声合成器を使用できる。

【００４５】

図１に戻り、演奏データ１はＭＩＤＩ音源９に渡され、ＭＩＤＩ音源９は演奏データをもとに楽音の生成を行う。この楽音は伴奏波形１０である。

【００４６】

歌声波形８と伴奏波形１０はともに同期を取りミキシングを行うミキシング部１１に渡される。

【００４７】

ミキシング部１１では、歌声波形８と伴奏波形１０との同期を取りそれぞれを重ね合わせて出力波形３として再生を行うことにより、演奏データ１をもとに伴奏を伴った歌声による音楽再生を行う。

【００４８】

ここで、ＭＩＤＩ音源９での楽音の再生はＭＩＤＩ制御部１２によりＭＩＤＩ制御データ１６に指示されているトラック又はチャンネルに対して、ミュートや音量の調節を行った上で再生が行われる。

【００４９】

ＭＩＤＩ制御データ１６は、歌詞付与部５において歌詞を付与する際にどのトラックに対して歌詞を付与するかを判別、設定するトラック選択部１３において選択されたトラック又はチャンネルの情報も反映され、ＭＩＤＩ音源９からの楽音と歌声生成部７が生成する歌声データを同時に再生する際に、歌声の対象となるトラック又はチャンネルに対して自動的にミュート又は音量の調整の処置を施すことができる。

【００５０】

また、これとは別にオペレータの指示により、任意のトラック又はチャンネルに対してもミュート又は音量の調整を施すことができる。

【００５１】

これらのＭＩＤＩ制御データ１６は演奏の対象となるＭＩＤＩデータと例えばファイル名が同じで拡張子が異なるなどの形で関連付けを持って保存することが可能である。

【００５２】

一般にＭＩＤＩ音源９は再生する楽音をwav形式などの波形データとして保存することも可能である。ミキシング部１１は歌声データとのミキシングを行う際に、この予め用意されたＭＩＤＩ楽音データの波形データと歌声データの波形を重ね合わせることによりミキシングを行うことも可能である。

【００５３】

ＤＴＭ（desk top music）等のシーケンサでは音声波形（wav形式）のデータを扱えるのは一般的である。上記のように音声波形としてまとめてしまえばＤＴＭ等のシーケンサにおいて、音声波形として取り込むことが可能になり、ＭＩＤＩの楽音とのミキシング処理自体もシーケンサにより行うことが可能である。

【００５４】

一般にＭＩＤＩ音源９はそのクロック等の違いにより音源の種類により再生される楽音が長い場合にわずかながらズレを生じることが知られている。ズレ補正部１４ではこのズレを補正するためにＭＩＤＩ音源９の種類にあわせてズレ補正データ１５内に予め用意された閾値を歌声生成部７において歌声を生成する際の時間データに対して掛け合わせることにより補正を行う。

【００５５】

このズレ補正データ１５は歌声生成部７が動作しているＣＰＵやＯＳ（operating system）などの環境とＭＩＤＩ音源の種類の組み合わせによって決まるが、それ以外にオペレータの指示によりこの閾値を変更することも可能である。

【００５６】

なお、歌声情報に関して、演奏データに歌詞が含まれている場合を説明したが、これには限られず、演奏データに歌詞が含まれない場合に任意の歌詞、例えば「ら」や「ぼん」等を自動生成し、又はオペレータにより入力し、歌詞の対象とする演奏データ（トラック、チャンネル）を、トラック選択部、歌詞付与部を介して選択して歌詞を割り振るようにしてもよい。

【００５７】

図５に図１に示す歌声合成装置の全体動作をフローチャートで示す。

【００５８】

先ずＭＩＤＩファイルの演奏データ１を入力する（ステップＳ１）。次に演奏データ１を解析し、楽譜データ４を作成する（ステップＳ２、Ｓ３）。次にオペレータに問い合わせオペレータの設定処理（例えば、歌詞の対象とするトラックやチャンネルの指定、ミュート又は音量調整すべきトラック又はチャンネルの指定、wavの作成指示、ＤＴＭへの取込指示等）を行う（ステップＳ４）。なおオ

ペレータが設定しなかった部分についてはデフォルトが後続処理で使用される。

【００５９】

次に、作成した楽譜データに基づき、歌詞を対象とするトラック又はチャンネルの演奏データに割り振って歌声情報６を作成する（ステップＳ５、Ｓ６）。

【００６０】

次に上述したタイミングのズレ補正閾値を取得し（ステップＳ７）、歌声生成部７において歌声情報６から歌声を生成する際の時間データに対して掛け合わせることにより補正を行って、音声波形（歌声波形８）を作成する。

【００６１】

次に、ＭＩＤＩ制御データ１６を参照して、ミュートすべきトラック、チャンネル又は音量調整すべきトラック、チャンネルがあるかチェックし（ステップ９）該当するＭＩＤＩトラック、チャンネルについては対応する処理をする（ステップＳ１０）。代表的には、歌詞の対象とした演奏データ（ＭＩＤＩトラック、チャンネル）は再生されないか、歌声に比べ小さな音量で再生されるよう音量調整処理される。

【００６２】

次に、ＭＩＤＩからwav形式の作成が指示されているかチェックする（ステップＳ１１）。指示されてなければ、ＭＩＤＩ再生をスタートさせ（ステップＳ１３）、歌声波形８と伴奏波形１０との同期を取りながらミキシングする（ステップＳ１７）。

【００６３】

ＭＩＤＩからwav形式の作成が指示されているときは、伴奏波形１０を作成した（ステップＳ１４）後、ＤＴＭへの取込が指示されているかチェックする（ステップＳ１５）。指示されていれば歌声波形８と共に伴奏波形１０をＤＴＭに引き渡す。指示されてなければ歌声波形８と伴奏波形１０を重ね合わせる（ステップＳ１６）。

【００６４】

ステップＳ１３又はＳ１６の後、Ｄ／Ａ変換器、アンプ、スピーカを含むサウンドシステム（図示せず）を介して歌声に伴奏の付いた音響信号を出力する（ス

テップＳ１７）。

【００６５】

なお、ステップＳ１２、Ｓ１３を通ってＳ１７に進む処理は、代表的には逐次的に実行される。すなわち、ＭＩＤＩの再生スタートを合図に、順次、リアルタイムでミキシングの実行とサウンドシステムによる音再生が行われる。これに対し、ステップＳ８からステップＳ１４、Ｓ１６を経てステップＳ１７に至る処理の場合、代表的には、いったん（予め）歌声と伴奏音の波形を作成し、重ね合わせてミキシングし、その結果を保存した後に、楽曲のサウンド再生要求に応じて音再生が行われる。

【００６６】

以上説明した歌声合成機能は例えば、ロボット装置に搭載される。

【００６７】

以下、一構成例として示す２足歩行タイプのロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、人間が行う基本的な動作を表出できるエンターテインメントロボットである。

【００６８】

図６に示すように、ロボット装置６０は、体幹部ユニット６２の所定の位置に頭部ユニット６３が連結されると共に、左右２つの腕部ユニット６４Ｒ／Ｌと、左右２つの脚部ユニット６５Ｒ／Ｌが連結されて構成されている（ただし、Ｒ及びＬの各々は、右及び左の各々を示す接尾辞である。以下において同じ。）。

【００６９】

このロボット装置６０が具備する関節自由度構成を図７に模式的に示す。頭部ユニット６３を支持する首関節は、首関節ヨー軸１０１と、首関節ピッチ軸１０２と、首関節ロール軸１０３という３自由度を有している。

【００７０】

また、上肢を構成する各々の腕部ユニット６４Ｒ／Ｌは、、肩関節ピッチ軸１０７と、肩関節ロール軸１０８と、上腕ヨー軸１０９と、肘関節ピッチ軸１１０と、前腕ヨー軸１１１と、手首関節ピッチ軸１１２と、手首関節ロール軸１１３

と、手部１１４とで構成される。手部１１４は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部１１４の動作は、ロボット装置６０の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は７自由度を有するとする。

【００７１】

また、体幹部ユニット６２は、体幹ピッチ軸１０４と、体幹ロール軸１０５と、体幹ヨー軸１０６という３自由度を有する。

【００７２】

また、下肢を構成する各々の脚部ユニット６５Ｒ／Ｌは、股関節ヨー軸１１５と、股関節ピッチ軸１１６と、股関節ロール軸１１７と、膝関節ピッチ軸１１８と、足首関節ピッチ軸１１９と、足首関節ロール軸１２０と、足部１２１とで構成される。本明細書中では、股関節ピッチ軸１１６と股関節ロール軸１１７の交点は、ロボット装置６０の股関節位置を定義する。人体の足部１２１は、実際には多関節・多自由度の足底を含んだ構造体であるが、ロボット装置６０の足底は、ゼロ自由度とする。したがって、各脚部は、６自由度で構成される。

【００７３】

以上を総括すれば、ロボット装置６０全体としては、合計で３＋７×２＋３＋６×２＝３２自由度を有することになる。ただし、エンターテインメント向けのロボット装置６０が必ずしも３２自由度に限定されるわけではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができることはいうまでもない。

【００７４】

上述したようなロボット装置６０がもつ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、２足歩行という不安定構造体に対して姿勢制御を行うことなどの要請から、アクチュエータは小型かつ軽量であることが好ましい。また、アクチュエータは、ギア直結型でかつサーボ制御系をワンチップ化してモータユニット内に搭載したタイプの小型ＡＣサーボ・アクチュエータで構成することがより好ましい。

【００７５】

図８には、ロボット装置６０の制御システム構成を模式的に示している。図８に示すように、制御システムは、ユーザ入力などに動的に反応して情緒判断や感情表現を司る思考制御モジュール２００と、アクチュエータ３５０の駆動などロボット装置６０の全身協調運動を制御する運動制御モジュール３００とで構成される。

【００７６】

思考制御モジュール２００は、情緒判断や感情表現に関する演算処理を実行するＣＰＵ（Central Processing Unit）２１１や、ＲＡＭ（Random Access Memory）２１２、ＲＯＭ（Read only Memory）２１３、及び、外部記憶装置（ハード・ディスク・ドライブなど）２１４で構成される、モジュール内で自己完結した処理を行うことができる、独立駆動型の情報処理装置である。

【００７７】

この思考制御モジュール２００は、画像入力装置２５１から入力される画像データや音声入力装置２５２から入力される音声データなど、外界からの刺激などに従って、ロボット装置６０の現在の感情や意思を決定する。ここで、画像入力装置２５１は、例えばＣＣＤ（Charge Coupled Device）カメラを複数備えており、また、音声入力装置２５２は、例えばマイクロホンを複数備えている。

【００７８】

また、思考制御モジュール２００は、意思決定に基づいた動作又は行動シーケンス、すなわち四肢の運動を実行するように、運動制御モジュール３００に対して指令を発行する。

【００７９】

一方の運動制御モジュール３００は、ロボット装置６０の全身協調運動を制御するＣＰＵ３１１や、ＲＡＭ３１２、ＲＯＭ３１３、及び外部記憶装置（ハード・ディスク・ドライブなど）３１４で構成される、モジュール内で自己完結した処理を行うことができる、独立駆動型の情報処理装置である。外部記憶装置３１４には、例えば、オフラインで算出された歩行パターンや目標とするＺＭＰ軌道、その他の行動計画を蓄積することができる。ここで、ＺＭＰとは、歩行中の床

反力によるモーメントがゼロとなる床面上の点のことであり、また、ＺＭＰ軌道とは、例えばロボット装置６０の歩行動作期間中にＺＭＰが動く軌跡を意味する。なお、ＺＭＰの概念並びにＺＭＰを歩行ロボットの安定度判別規範に適用する点については、Miomir Vukobratovic 著“LEGGED LOCOMOTION ROBOTS”（加藤一郎外著『歩行ロボットと人工の足』（日刊工業新聞社））に記載されている。

【００８０】

運動制御モジュール３００には、図８に示したロボット装置６０の全身に分散するそれぞれの関節自由度を実現するアクチュエータ３５０、体幹部ユニット６２の姿勢や傾斜を計測する姿勢センサ３５１、左右の足底の離床又は着床を検出する接地確認センサ３５２，３５３、バッテリなどの電源を管理する電源制御装置３５４などの各種の装置が、バス・インターフェース（Ｉ／Ｆ）３０１経由で接続されている。ここで、姿勢センサ３５１は、例えば加速度センサとジャイロ・センサの組み合わせによって構成され、接地確認センサ３５２，３５３は、近接センサ又はマイクロ・スイッチなどで構成される。

【００８１】

思考制御モジュール２００と運動制御モジュール３００は、共通のプラットフォーム上で構築され、両者間はバス・インターフェース２０１，３０１を介して相互接続されている。

【００８２】

運動制御モジュール３００では、思考制御モジュール２００から指示された行動を体現すべく、各アクチュエータ３５０による全身協調運動を制御する。すなわち、ＣＰＵ３１１は、思考制御モジュール２００から指示された行動に応じた動作パターンを外部記憶装置３１４から取り出し、又は、内部的に動作パターンを生成する。そして、ＣＰＵ３１１は、指定された動作パターンに従って、足部運動、ＺＭＰ軌道、体幹運動、上肢運動、腰部水平位置及び高さなどを設定するとともに、これらの設定内容に従った動作を指示する指令値を各アクチュエータ３５０に転送する。

【００８３】

また、ＣＰＵ３１１は、姿勢センサ３５１の出力信号によりロボット装置６０

の体幹部ユニット６２の姿勢や傾きを検出するとともに、各接地確認センサ３５２，３５３の出力信号により各脚部ユニット６５Ｒ／Ｌが遊脚又は立脚のいずれの状態であるかを検出することによって、ロボット装置６０の全身協調運動を適応的に制御することができる。

【００８４】

また、ＣＰＵ３１１は、ＺＭＰ位置が常にＺＭＰ安定領域の中心に向かうように、ロボット装置６０の姿勢や動作を制御する。

【００８５】

さらに、運動制御モジュール３００は、思考制御モジュール２００において決定された意思通りの行動がどの程度発現されたか、すなわち処理の状況を、思考制御モジュール２００に返すようになっている。

【００８６】

このようにしてロボット装置６０は、制御プログラムに基づいて自己及び周囲の状況を判断し、自律的に行動することができる。

【００８７】

このロボット装置６０において、上述した歌声合成機能をインプリメントしたプログラム（データを含む）は例えば思考制御モジュール２００のＲＯＭ２１３に置かれる。この場合、歌声合成プログラムの実行は思考制御モジュール２００のＣＰＵ２１１により行われる。

【００８８】

このようなロボット装置に上記歌声合成機能を組み込むことにより、伴奏に合わせて歌うロボットとしての表現能力が新たに獲得され、エンターテインメント性が広がり、人間との親密性が深められる。

【００８９】

なお、本発明は上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。

【００９０】

例えば、本件出願人が先に提案した特願２００２－７３３８５の明細書及び図面に記載の音声合成方法及び装置等に用いられる歌声合成部及び波形生成部に対

応した歌声生成部７に使用可能な歌声情報を例示しているが、この他種々の歌声生成部を用いることができ、この場合、各種の歌声生成部によって歌声生成に必要とされる情報を含むような歌声情報を、上記演奏データから生成するようにすればよいことは勿論である。また、演奏データは、ＭＩＤＩデータに限定されず、種々の規格の演奏データを使用可能である。

【００９１】

【発明の効果】

以上詳細に説明したように、本発明に係る歌声合成方法及び装置によれば、演奏データを音の高さ、長さ、歌詞の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、ＭＩＤＩデータに代表されるような演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏として歌詞を歌い上げることができる。したがって、従来、楽器の音のみにより表現していた音楽の作成や再生において特別な情報を加えることがなく歌声の合成を行うことによりその音楽表現は格段に向上する。

【００９２】

また、本発明に係るプログラムは、本発明の歌声合成機能をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【００９３】

本発明に係るプログラム及び記録媒体によれば、演奏データを音の高さ、長さ、歌詞の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏とした歌唱が可能となる。

【００９４】

また、本発明に係るロボット装置は本発明の歌声合成機能を実現する。すなわち、本発明のロボット装置によれば、供給された入力情報に基づいて動作を行う自律型のロボット装置において、入力された演奏データを音の高さ、長さ、歌詞

の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、ＭＩＤＩデータに代表されるような演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏として歌詞を歌い上げることができる。したがって、ロボット装置の表現能力が向上し、エンターテインメント性を高めることができると共に、人間との親密性を深めることができる。

【図面の簡単な説明】

【図１】

本実施の形態における歌声合成装置のシステム構成を説明するブロック図である。

【図２】

解析結果の楽譜情報の例を示す図である。

【図３】

歌声情報の例を示す図である。

【図４】

歌声生成部の構成例を説明するブロック図である。

【図５】

本実施の形態における歌声合成装置の動作を説明するフローチャートである。

【図６】

本実施の形態におけるロボット装置の外観構成を示す斜視図である。

【図７】

同ロボット装置の自由度構成モデルを模式的に示す図である。

【図８】

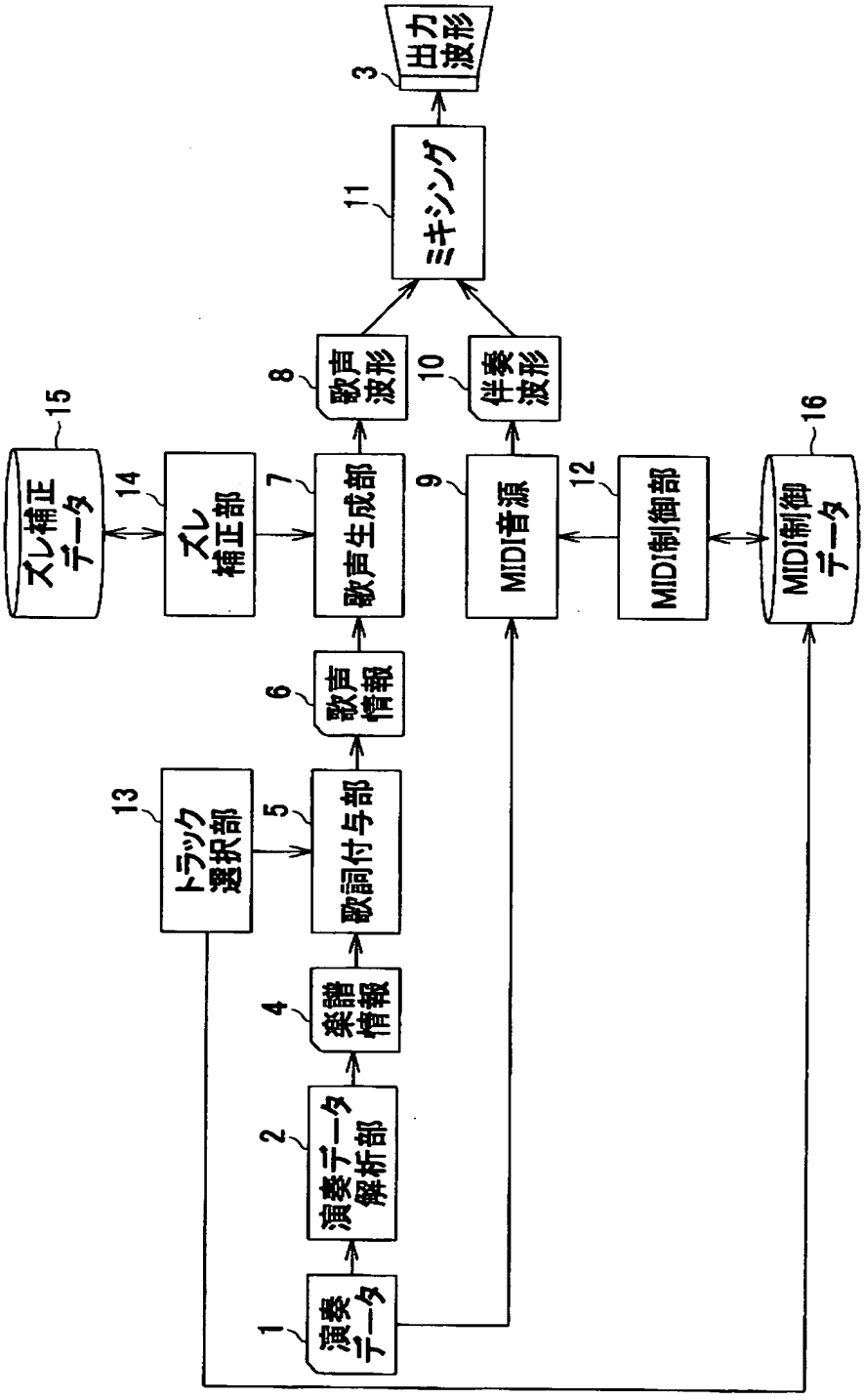
同ロボット装置のシステム構成を示すブロック図である。

【符号の説明】

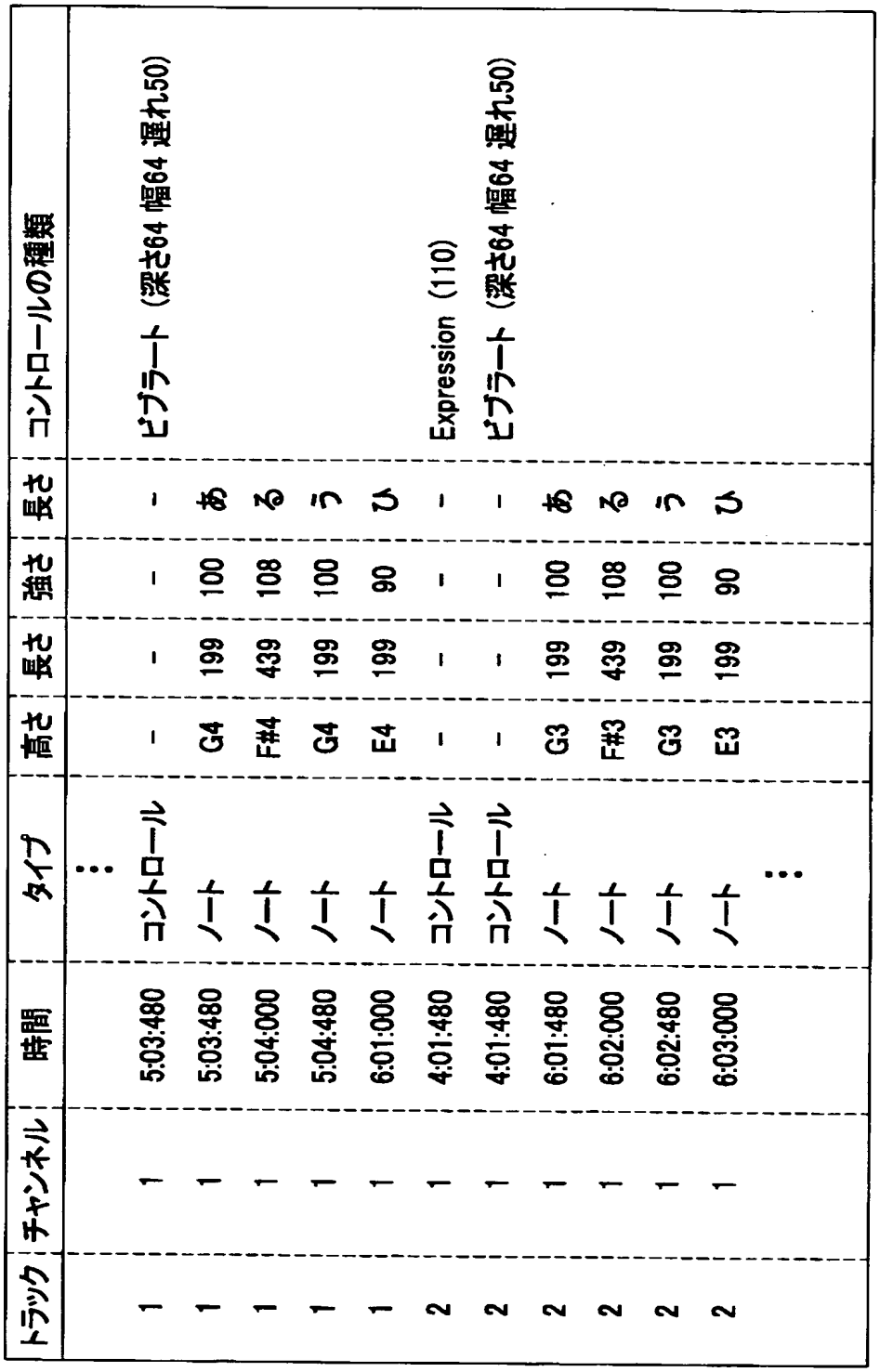
２　演奏データ解析部、５　歌詞付与部、７　歌声生成部、９　ＭＩＤＩ音源１１　ミキシング部、１２　ＭＩＤＩ制御部、６０　ロボット装置、２１１　ＣＰＵ、２１３　ＲＯＭ

【書類名】　図面

【図１】


出力波形
3
ミキシング
11
歌声波形
8
伴奏波形
10
歌声生成部
7
MIDI音源
9
ズレ補正データ
15
ズレ補正部
14
MIDI制御部
12
MIDI制御データ
16
歌声情報
6
歌詞付与部
5
トラック選択部
13
楽譜情報
4
演奏データ解析部
2
演奏データ
1

【図２】

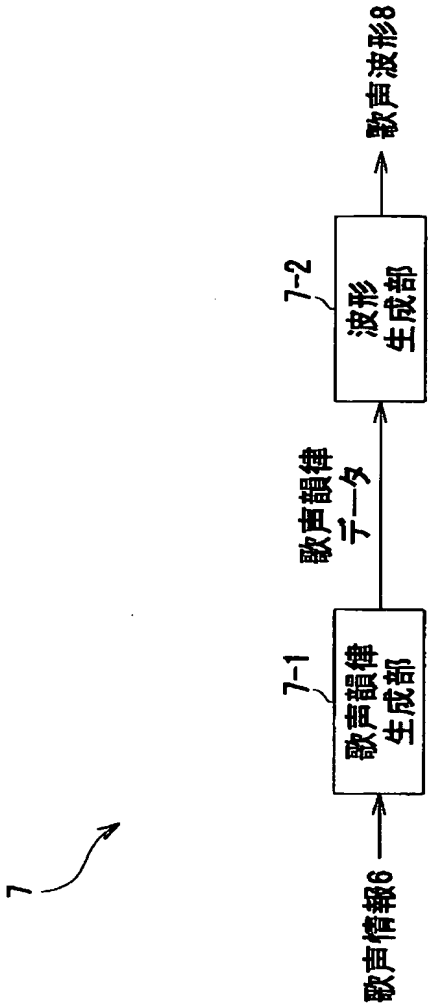


トラック	チャンネル	時間	タイプ	高さ	長さ	強さ	長さ	コントロールの種類
			・・・					
1	1	5:03:480	コントロール	-	-	-	-	ピブラート（深さ64 幅64 遅れ50）
1	1	5:03:480	ノート	G4	199	100	あ	
1	1	5:04:000	ノート	F#4	439	108	る	
1	1	5:04:480	ノート	G4	199	100	う	
1	1	6:01:000	ノート	E4	199	90	ひ	
2	1	4:01:480	コントロール	-	-	-	-	Expression（110）
2	1	4:01:480	コントロール	-	-	-	-	ピブラート（深さ64 幅64 遅れ50）
2	1	6:01:480	ノート	G3	199	100	あ	
2	1	6:02:000	ノート	F#3	439	108	る	
2	1	6:02:480	ノート	G3	199	100	う	
2	1	6:03:000	ノート	E3	199	90	ひ	
			・・・					

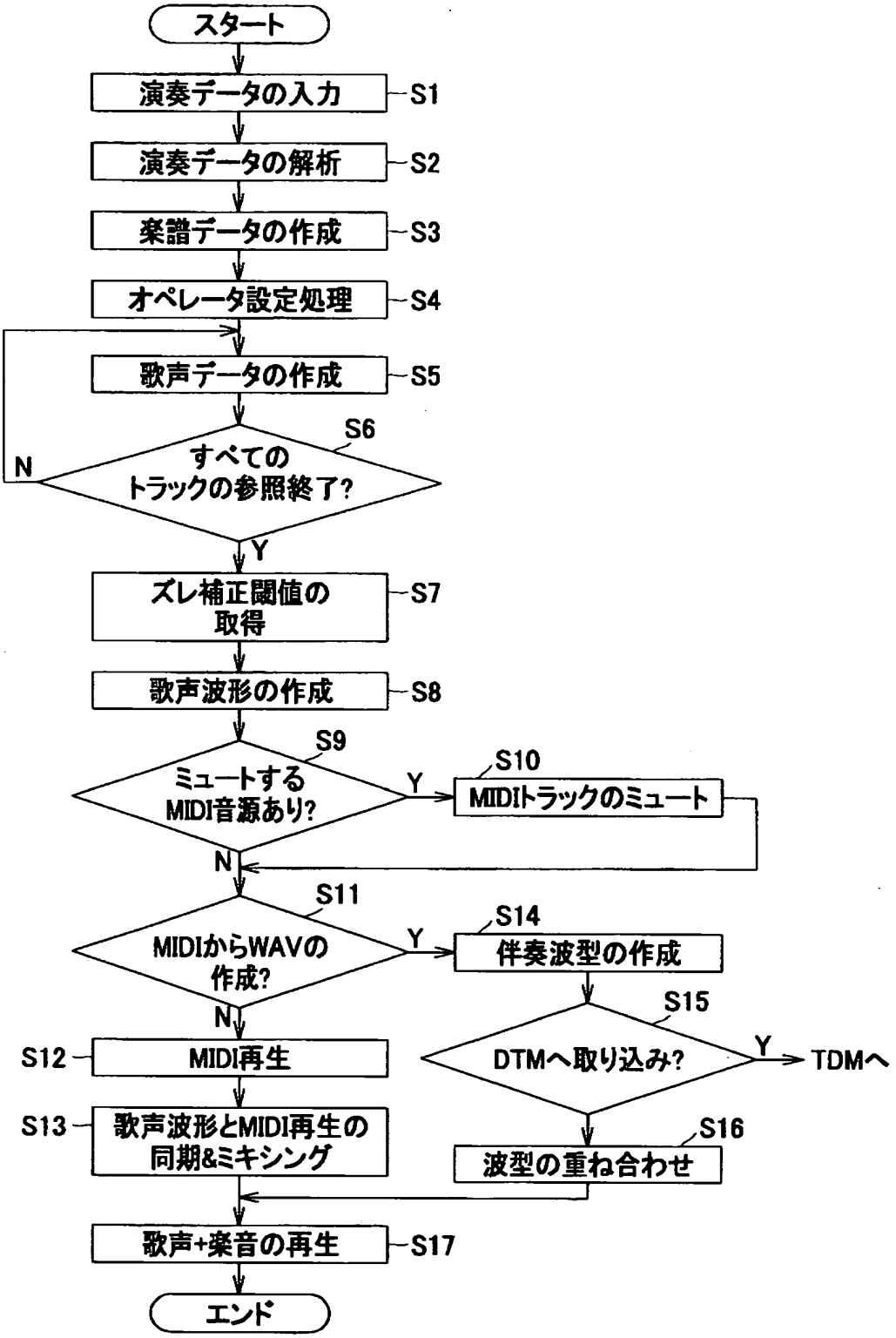
【図３】

¥song¥　　←―― 歌声データの開始
¥PP,T10673075¥　　←―― 10673075μ secの休み
¥tdyna 110 649075¥　　←―― 先頭から10673075μ secの全体の強さ
¥fine-100¥　　←―― ピッチの微調整(MIDIのファインチューンに同じ)
¥vibrato NRPN_dep=64¥　　←―― ビブラート
¥vibrato NRPN_del=50¥
¥vibrato NRPN_rat=64¥
¥dyna 100¥　　←―― 音ごとの強弱
¥G4,T288461¥あ　　←―― G4の高さの音で、長さが288461μsec、歌詞は「あ」
¥dyna 108¥
¥Gb4,T288462¥る
¥dyna 100¥
¥G4,T288461¥う
¥dyna 90¥
¥E4,T219592¥ひ
¥PP,T1222716¥
¥dyna 100¥
¥E4,T144231¥も
¥dyna 98¥
¥E4,T144230¥り
⋮

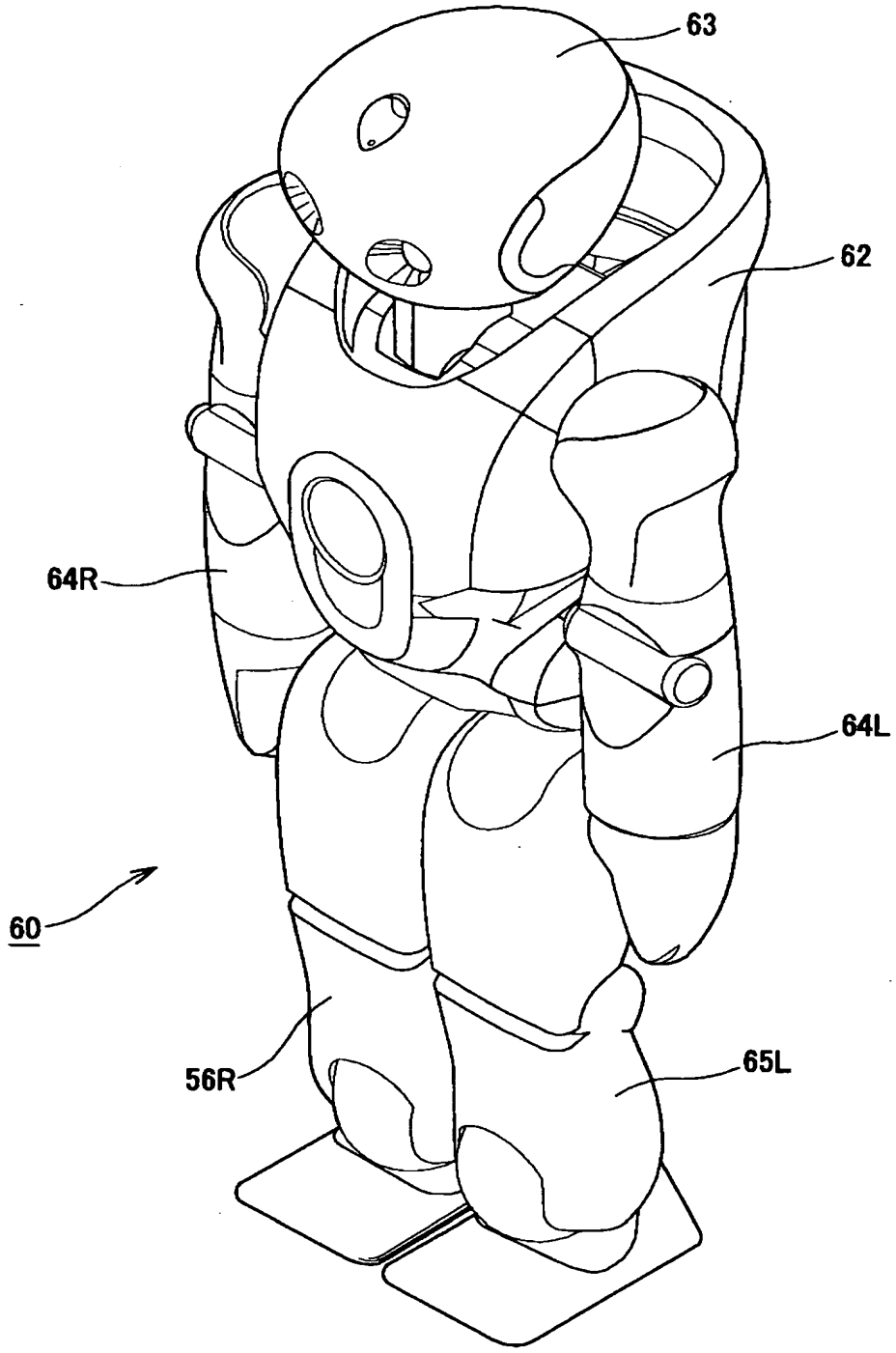
【図４】



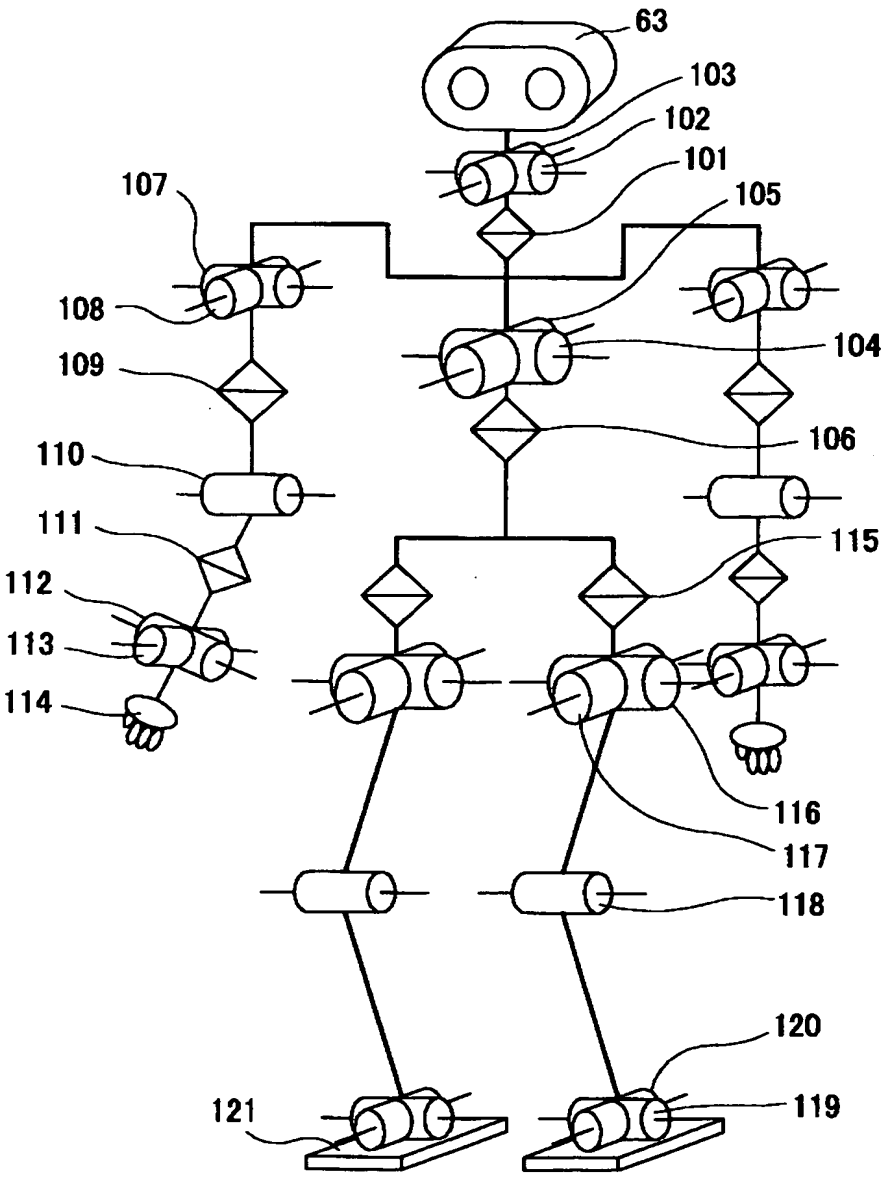
【図５】


スタート
演奏データの入力
S1
演奏データの解析
S2
楽譜データの作成
S3
オペレータ設定処理
S4
歌声データの作成
S5
S6
すべての
トラックの参照終了?
N
Y
ズレ補正閾値の
取得
S7
歌声波形の作成
S8
S9
ミュートする
MIDI音源あり?
Y
S10
MIDIトラックのミュート
N
S11
MIDIからWAVの
作成?
Y
S14
伴奏波型の作成
N
S15
DTMへ取り込み?
Y
TDMへ
S12
MIDI再生
S13
歌声波形とMIDI再生の
同期&ミキシング
S16
波型の重ね合わせ
歌声+楽音の再生
S17
エンド

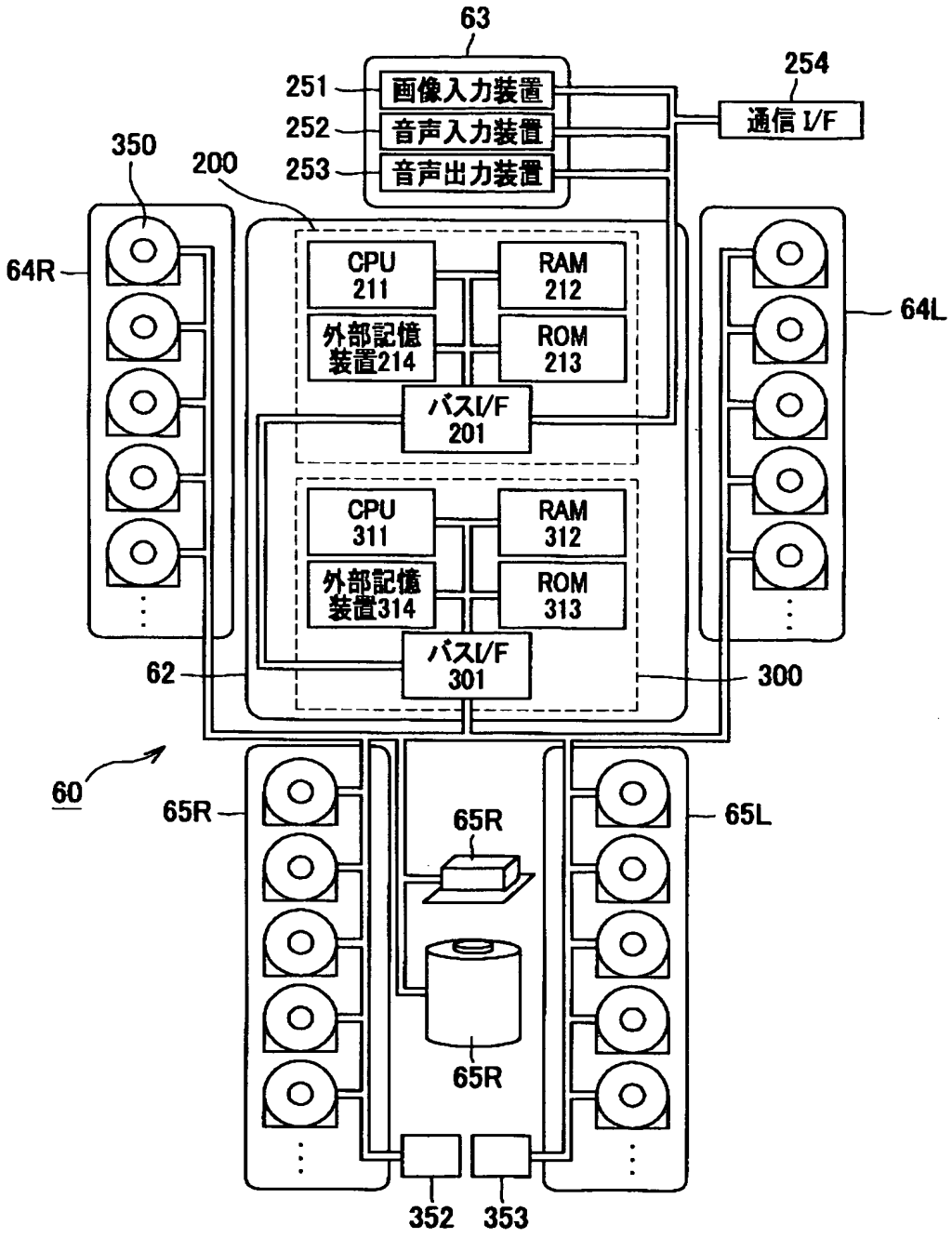
【図６】



【図７】


63
103
102
101
107
105
108
104
109
106
110
111
115
112
113
114
116
117
118
120
121
119

【図８】


63
251
画像入力装置
252
音声入力装置
253
音声出力装置
254
通信 I/F
350
200
64R
64L
CPU
211
RAM
212
外部記憶
装置214
ROM
213
バスI/F
201
CPU
311
RAM
312
外部記憶
装置314
ROM
313
バスI/F
301
62
300
60
65R
65R
65L
65R
352
353

【書類名】　要約書

【要約】

【課題】　ＭＩＤＩデータ等の演奏データを活用して歌声を合成する。

【解決手段】　入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する（Ｓ２、Ｓ３）。解析された音楽情報から歌詞を音列に付与して歌声データを作成する（Ｓ５、Ｓ６）。歌声データから歌声の音声波形を作成する（Ｓ７、Ｓ８）。入力された演奏データから楽音の波形を作成する（Ｓ１４）。歌声に使用した演奏データは楽音の再生に使用しないか再生の音量を抑制することが好ましい。

【選択図】　図５

特願２００３－０７９１５１

出　願　人　履　歴　情　報

識別番号　［０００００２１８５］

１．変更年月日　１９９０年　８月３０日
［変更理由］　新規登録
住　所　東京都品川区北品川６丁目７番３５号
氏　名　ソニー株式会社